METHOD FOR SINGLE NUCLEOTIDE POLYMORPHISM DETECTION

Introduction

This invention was supported in part by funds from the U.S. government (NIH Grant No. AI46669) and the U.S. government may therefore have certain rights in the invention.

Field of the Invention

The present invention provides an improved, cost-effective, high-throughput assay useful in single nucleotide polymorphism (SNP) detection. In this method, the amplification refractory mutation (ARMS) assay has been modified to use hairpin shaped primers targeted to the mutation site. This method is extremely useful in efficiently identifying SNPs responsible for drug resistance of infective organisms.

Background of the Invention

The acquisition of mutations and single nucleotide polymorphisms (SNPs) in target genes is a major cause of the development of drug resistance, variations in immunity, and changes in phenotype of infectious organisms. For example, most causes of drug-resistance in *M. tuberculosis* appear to be the result of SNPs in particular target genes. However, each SNP occurs at relatively low frequency. Therefore, in the absence of large sequencing studies, it has been difficult to establish statistically valid associations between individual SNPs and resistance to a particular drug. In the case of resistance to the antibiotic isoniazid (INH), only mutations in codon 315 of the *katG* gene occur with sufficient frequency. The danger of reaching conclusions after analyzing only a limited number of *M. tuberculosis*

isolates is illustrated by the case of SNPs in codon 463 of *katG*, and codons 269, and 312 of the *kasA* gene, which were originally identified as resistance associated, but were later shown to be common in INH-susceptible isolates (Abate et al. Eur. J. Clin. Microbiol. Infect. Dis. 2001 20:329-33; Alland et al. J. Bacteriol. 2003 185:3392-9; Lee et al. Antimicrob. Agents Chemother. 1999 43:2087-9; Saint-Joanis et al. Biochem. J. 338 (Pt 3):753-60; Sreevatsan et al. Proc. Natl Acad. Sci. USA 1997 94:9869-74). Accordingly, a comprehensive understanding of the frequency and position of mutations is needed to design genetic assays to detect drug resistance, perform epidemiological investigations identify potential drug targets, identify strains with different immunogenicity, and identify strains with differences in virulence or other phenotype.

SNPs are also a major cause of disease in human and other higher organisms as well as a major determinant of susceptibility to infection. SNPs in certain genes can also determine human response to therapeutic drugs and can also have important effects on susceptibilities to drug related toxicities.

The amplification refractory mutation (ARMS) assay is used routinely to identify SNPs. In this assay two allele-specific primers are designed that are identical except for their 3′ end nucleotide. At the last 3′ nucleotide, one primer is designed to be complementary to the wild-type sequence, while the other primer is complementary to the mutant sequence. Two PCR reactions are performed, one with each of the two 3′ variable primers. Under usual circumstances, the reaction containing the perfectly complementary primer is more efficient than the reaction containing the 3-prime mismatched primer. Thus, the reaction with the complementary primer will contain more PCR product at the end of PCR amplification. This result identifies the correct allele of the target sequence (Newton et al. Nucleic Acids Res. 1989 17:2503-16; Sommer et al. Mayo

Clin. Proc. 1989 64:1361-72; Wu et al. Proc. Natl Acad. Sci. USA 1989 86:2757-60).

The ARMS technique can be improved by using real-time PCR (Germer et al. Genome Res. 2000 10:258-66). In real-time PCR, fluorescent techniques measure amplicon synthesis at the annealing or extension segment of each PCR cycle (Bassler et al. Appl. Environ. Microbiol. 1995 61:3724-3728; Livak et al. PCR Methods Appl. 1995 4:357-362). When the fluorescence intensity is plotted at each PCR cycle, successful real-time PCR reactions generate a characteristic rising curve. The point at which the fluorescence of a real-time PCR reaction becomes detectable is referred to as the "threshold cycle".

In the ARMS method, the well with the perfect primer-template match amplifies more efficiently and has a shorter threshold cycle than the well with the 3′ mismatched primer. Therefore in principle, in this method, the correct allele for each SNP can be determined by comparing the relative threshold cycle values for the each of the paired assay wells.

However, while simple in principle, a number of problems with this method make it less reliable in practice. For example, one problem is that many 3′ mismatches do not destabilize the primer-template hybrid sufficiently enough to delay the cycle threshold, thus making it impossible to assay for SNPs. Another major problem is that PCR amplification of chromosomal DNA under conditions that permit resolution of each SNP is often too inefficient to result in a detectable signal.

In an effort to resolve some of these problems, most 3′ mismatch techniques "preamplify" the target in a nested or hemi-nested protocol. However, nested PCR protocols entail post-PCR manipulations to prepare for the nested reactions.

The present invention provides an improved, cost-effective, high-throughput assay useful in SNP detection.

Summary of the Invention

An object of the present invention is to provide a method for detecting single nucleotide polymorphisms in an organism via a modified amplification refractory mutation assay that utilizes a hairpin shaped primer pair that discriminates between different alleles by situating its 3′ nucleotide at the location of a SNP.

Another object of the present invention is to provide an assay kit for detection of a single nucleotide polymorphism comprising a hairpin shaped primer pair that discriminates between different alleles by situating its 3′ nucleotide at the location of a SNP.

Brief Description of the Figures

Figures 1A and 1B provide schematic representations of an assay of the present invention using hairpin (HP) shaped primers. Two reactions with the same DNA target are performed in parallel. Figure 1A provides a schematic of the HP sequence being fully complementary to the target DNA sequence. Figure 1B provides a schematic wherein the HP sequence is complementary to the target DNA sequence except for the last 3′ nucleotide of the primer which is complementary to an alternate allelic sequence. As shown in Figure 1C, the real-time PCR fluorescence curve develops more rapidly in the well where the HP sequence is fully complementary to the target DNA sequence (earlier Ct), indicating the presence of allele A.

Figure 2A through 2D provide a comparison of HP (Figure 2A), linear primer (LP; Figure 2B), linearized-HP tail (LHP; Figure 2C) and substituted extended-LHP tail (ELHP; Figure 2D) with and without secondary mutations. Ct for each paired reactions are shown as a single point. The X axis denotes the Ct of the reaction with allele "A" primers and the Y axis denotes the Ct of the reaction with allele "B" primers.

Values above the diagonal line (marking the place where Ct(B)/Ct(A) > 1) indicate that an A allele SNP should be present, values below the diagonal indicate that a B allele SNP should be present. Primers with secondary mutations are represented by diamonds, and primers without secondary mutations are represented by triangles. Closed symbols correspond to successful assays; open symbols indicate erroneous or indeterminate (ΔCt<5) assays.

Figure 3A-3C show the primers (Figure 3A), their secondary structure (Figure 3B) and results (Figure 3C) of detecting four alleles at a single codon using the HP assay. Assays were designed to detect mutations S315I (AGC→ATC), S315N (AGC→AAC) and S315T (AGC→ACC) in the *katG* gene. Figure 3A shows the DNA sequence of the target (SEQ ID NO:120) and HP primers FHP $katG^{S315T}$ (SEQ ID NO:121), FHP $katG^{S315N}$ (SEQ ID NO:122), FHP $katG^{S315I}$ (SEQ ID NO:123) and FHP $katG^{S315}$ (SEQ ID NO:124). Secondary mutations and SNPs are shown in bold capitals and 5'-end tails are underlined. The location of the constant primer is indicated by lower-case bolding in the target (SEQ ID NO:120). Figure 3B shows predicted secondary structures for each of the HP primers at 60°C, 2 mM $MgCl_2$ (annealing conditions). Figure 3C shows real time PCR results using chromosomal DNA from H37Rv as WT control (*katG*315; filled square), and the MUT isolates I-524 (*katG*S315I; open diamond), M-5036 (*katG*S315N; filled circle) and M-5153 (*katG*S315T; filled triangle). An earlier Ct was observed for each matched reaction, indicating the correct allele.

Detailed Description of the Invention

The present invention provides low-cost, high-throughput methods and assay kits for detection of single nucleotide polymorphisms (SNPs) in an organism. The methods and kits of the present invention enable analysis of a large number of

isolates, thereby providing a means for comprehensive understanding of the frequency and position of mutations in an organism. In the present invention, the amplification refractory mutation or ARMS assay has been modified to utilize hairpin shaped primers. The hairpin shaped primers used in the methods and assay kits of the present invention discriminate between different alleles by situating their 3′ nucleotide at the location of the SNP.

For purposes of the present invention, by the term organism it is meant to be inclusive of any organism with nucleic acid sequences including viruses, prokaryotes, and eukaryotes, including humans.

By hairpin shaped primer, as used herein, it is meant an oligonucleotide with a 5′ tail that hybridizes to its 3′ end. Accordingly, hairpin primers used in the present invention form a stem-and-loop structure. In a preferred embodiment, the hairpin shaped primers have a melting point (Tm) of about 5°C to about 10°C above the Tm of the primer-target hybrid. However, hairpin shaped primers with tails that exhibit a Tm outside of this preferred range may also be used. The length or size of the primer is selected based upon the desired annealing temperature in the PCR reaction. A preferred size of the primer is about 15 to about 25 bases. In general, the primers are comprised of DNA. However RNA, PNA, and modified nucleotides can also be used in the primers of the present invention.

Use of hairpin shaped primers in the method of the present invention decreases non-specific primer target hybridizations observed with the traditional ARMS assay during the early stages of PCR, thus greatly simplifying assay design and interpretation. Furthermore, it has recently been reported that hairpin primers are less likely to form primer dimers in PCR (Nazarenko et al. Nucleic Acids Res. 2002 30(9):e37). In addition, in contrast to Taqman probes and molecular beacons, the hairpin shaped primers used in the present invention do not

require fluorophore labeling for detection in simple reactions thereby dramatically lowering assay costs. Hairpin shaped primers of the present invention can be labeled with a fluorophore, however, to permit multiplexed reactions monitored by real-time PCR.

To produce hairpin shaped primers for use in the SNP detection methods and assay kits of the present invention, pairs of primers that varied at the last 3′ nucleotide were first prepared. As in standard 3′ primer mismatch PCR, one primer was designed to be complementary to the wild-type sequence, while the other primer was designed to be complementary to the mutant sequence. Additional modifications were then made at the 5′ end of each primer so that it was reverse-complementary to the first 5-8 nucleotides of its 3′ end, conferring to it the ability to form a "hairpin". The primers were designed so that the melting temperature of the hairpin stems was approximately six degrees above the melting temperature of the primers to their complementary targets. Additional mutations were occasionally necessary in the "loop" of the primers to achieve the desired secondary structure and melting temperature. These primers formed a uni-molecular stem-and-loop structure in solution, but hybridized to complementary DNA targets to form linear bimolecular complexes below the melting temperature of the primers to their targets. It was found that this type of "hairpin" primer configuration greatly decreased the amplification efficiency of targets that were mismatched for a SNP at the 3′ nucleotide of the primer. This is believed to be due to the propensity of a hairpin primer to disassociate from a mismatched target in preference to forming a stable uni-molecular stem-and-loop structure. In contrast, hairpin primers that were perfectly matched to their targets formed more stable bimolecular primer-target hybrids. This enabled perfectly matched

hairpin primers to prime PCR reactions with amplification efficiencies similar to linear PCR primers. Mismatched hairpin primers usually delayed amplification by 5 to 18 PCR cycles. To increase amplification efficiency, the size of the amplicons was substantially reduced from typical size of approximately 100 to 300 bases. Most assays were designed to produce amplicons that are only a few base pairs longer than the sum of both PCR primers. By combining hairpin primers with these shorter amplicons, the success rate of each newly designed SNP assay has been dramatically improved.

Assays using the hairpin shaped primers are very easy to perform and analyze. A schematics of setting forth the principles of the assay are depicted in Figures 1A and 1B.

In a preferred embodiment, real-time PCR techniques are used for monitoring. For example, for a high-throughput screening assay, primer pairs (constant and hairpin shaped primer) can be added and dried in 384-well plates. By constant primer it is meant a linear or hairpin reverse primer that is conserved in both SNP alleles. The appropriate reaction mixes containing SYBR green dye, dNTPs, PCR buffer, water, $MgCl_2$ and Taq polymerase, optimized for the mutation are then added to the primer-containing wells, the plate is sealed, and real-time PCR is performed, for example, on an ABI 7900 real-time PCR machine or other real-time PCR instrument for 30 to 50 thermal-cycles. A touchdown protocol is usually incorporated as well. The threshold cycle is automatically calculated and the correct SNP allele is identified by comparing the threshold cycle of each paired well. The results can also be examined graphically to identify problematic results.

The specificity of real-time PCR reactions in this embodiment is high because all reactions are performed in sealed plates. These plates are never opened, therefore

false positive results caused by cross-contamination of amplicons do not occur.

In another assay embodiment, PCR is performed without real-time monitoring and amplicon production is measured at the completion of the PCR reaction.

The assays of the present invention can be performed at relatively low cost as compared to other methods for SNP detection. Amplicon generation by the method of the present invention can be detected by adding inexpensive SYBR green dye to the PCR mix. Accordingly, the assay does not require use of expensive fluorescent-labeled primers or probes. Costs are further decreased because reaction volumes can be as low as four microliters. In addition, the 384-well format of real-time PCR machines such as the ABI 7900 enables very high throughput.

Further, as will be understood by one of skill in the art upon reading the instant application, the SNP detection assay of the present invention can be routinely adapted for use of other amplification methods including, but not limited to, TMA and SDA.

The present invention also provides assay kits for detecting a single nucleotide polymorphism in an organism. Kits of the present invention comprise a hairpin shaped primer that discriminates between different alleles by situating its 3′ nucleotide at the location of a single nucleotide polymorphism. Kits also preferably comprise additional ingredients for use in the amplification method to be used as well as detection of the generated amplicons. For example, for amplification by PCR, the assay kits of the present invention may further comprise SYBR green dye and components for the PCR mixture, as well as additional primers labeled with a fluorophore.

The ability of the method of the present invention to detect SNPs rapidly and with high sensitivity was

demonstrated in *Mycobacterium tuberculosis* and other organisms. In these experiments, amplification refractory mutation (ARMS) SNP assays were modified by converting the SNP-detecting linear primers (LPs) in the ARMS assay to hairpin-shaped primers (HPs) through the addition of a 5' tail complementary to the 3'-end of the LP. The improved ability of these primers to detect SNPs in *M. tuberculosis* was compared in a real time PCR reaction using SYBR-I green dye. LPs resulted in incorrect or indeterminate allele designation for six of the thirteen SNP alleles tested in seven different SNP assays, while HPs determined the correct SNP in all cases. The cycle threshold differences (ΔCt) were also compared between the reactions containing primer-template matches and the reactions containing primer-template mismatches (where a larger ΔCt indicates a more robust assay). The use of HPs dramatically improved the mean ΔCt values for the SNP assays (7.6 for LPs and 11.2 for HPs).

Further, ninety-eight different HP assays were designed for SNPs previously associated with resistance to the antibiotic isoniazid to test the large-scale utility of the HP approach. Assay design was successful in 72.4%, 83.7%, 88.8% and 92.9% of the assays after one to four rounds of assay design respectively. Thus, the HP SNP assays of the present invention provide a simple, sensitive, robust, and inexpensive technique for SNP detection.

The following nonlimiting examples are provided to further illustrate the present invention.

EXAMPLE

Example 1: *M. tuberculosis* isolates and chromosomal DNA extraction

Chromosomal DNA was extracted using the CTAB method as described by van Embden et al. (J. Clin. Microbiol. 1993

31:406-9). The reference strain H37Rv was used as a wild-type (WT) control. Clinical isolates I-524, M-5036 and M-5455 were used as mutant (MUT) controls for *kat*GS315I, *kat*GS315N and *kat*GS315T, respectively. For the remaining SNPs tested, the isolates used are indicated in Table 1.

Table 1: Sequences of Primers Used

SNP	LP	HP	LHP	ELHP	Constant Primer	Amp. lng (bp)[2]	MUT Isolate
katG315[1]	catacgTcctcgatgccgc (SEQ ID NO:1)	GCGGC (SEQ ID NO:29)	CGCCC (SEQ ID NO:57)	ATATACGCCC (SEQ ID NO:85)	ccggtaaggac gcgatcac (SEQ ID NO:113)	40	M-5455
katGS315T[1]	catacgTcctcgatgccgG (SEQ ID NO:2)	CCGGC (SEQ ID NO:30)	CGCCC (SEQ ID NO:58)	ATATACGCCC (SEQ ID NO:86)			
katG315[1]	catacgacctcgatgccgc (SEQ ID NO:3)	GCGGC (SEQ ID NO:31)	CGCCC (SEQ ID NO:59)	ATATACGCCC (SEQ ID NO:87)			
katGS315T[1]	catacgacctcgatgccgG (SEQ ID NO:4)	CCGGC (SEQ ID NO:32)	CGCCC (SEQ ID NO:60)	ATATACGCCC (SEQ ID NO:88)			
katG485	cgtcCtcgttccgtgg (SEQ ID NO:5)	CCACGG (SEQ ID NO:33)	GCTCGC (SEQ ID NO:61)	GATCGCGCTCGC (SEQ ID NO:89)	aggcggatgcg acca (SEQ ID NO:114)	58	N/A[3]
katGG485V	cgtcCtcgttccgtgT (SEQ ID NO:6)	ACACGG (SEQ ID NO:34)	TCTCGC (SEQ ID NO:62)	GATCGCTCTCGC (SEQ ID NO:90)			
katG485	cgtcgtcgttccgtgg (SEQ ID NO:7)	CCACGG (SEQ ID NO:35)	GCTCGC (SEQ ID NO:63)	GATCGCGCTCGC (SEQ ID NO:91)			
katGG485V	cgtcgtcgttccgtgT (SEQ ID NO:8)	ACACGG (SEQ ID NO:36)	TCTCGC (SEQ ID NO:64)	GATCGCTCTCGC (SEQ ID NO:92)			
kasA269	cgattCctgggtgccg (SEQ ID NO:9)	CGGCA (SEQ ID NO:37)	GCGGA (SEQ ID NO:65)	GCGGAGCGGA (SEQ ID NO:93)	aaaggcgtccg aggtgatac (SEQ ID NO:115)	36	M-5279
kasAG269S	cgattCctgggtgccA (SEQ ID NO:10)	TGGCA (SEQ ID NO:38)	ACGGA (SEQ ID NO:66)	GCGGAACGGA (SEQ ID NO:94)			
kasA269	cgattgctgggtgccg (SEQ ID NO11)	CGGCA (SEQ ID NO:39)	GCGGA (SEQ ID NO:67)	GCGGAGCGGA (SEQ ID NO:95)			

kasAG269S	cgattgctgggtgcc**A** (SEQ ID NO:12)	TGGCA (SEQ ID NO:40)	ACGGA (SEQ ID NO:68)	GCGGAACGGA (SEQ ID NO:96)			
inhA21	atcatc**T**ccgactcgtcgat (SEQ ID NO:13)	ATCGACG (SEQ ID NO:41)	ATGCACC (SEQ ID NO:69)	ACGAGCAAATGCACC (SEQ ID NO:97)	gctacccgtgc gatgtgaa (SEQ ID NO:116)	39	M-5502
inhAI21T	atcatc**T**ccgactcgtcga**C** (SEQ ID NO:14)	GTCGACG (SEQ ID NO:42)	CTGCACC (SEQ ID NO:70)	ACGAGCAACTGCACC (SEQ ID NO:98)			
inhA21	aatcatcaccgactcgtcgat (SEQ ID NO:15)	ATCGACG (SEQ ID NO:43)	ATGCACC (SEQ ID NO:71)	ACGAGCAAATGCACC (SEQ ID NO:99)			
inhAI21T	aatcatcaccgactcgtcga**C** (SEQ ID NO:16)	GTCGACG (SEQ ID NO:44)	CTGCACC (SEQ ID NO:72)	ACGAGCAACTGCACC (SEQ ID NO:100)			
inhA94	ggca**A**gaacccaatcga (SEQ ID NO:17)	TCGATTGGA (SEQ ID NO:45)	TGGAATGCA (SEQ ID NO:73)	GAGGCAAGTGGAATGCA (SEQ ID NO:101)	cgggcaacaag ctcgac (SEQ ID NO:117)	46	M-5041
inhAS94A	ggca**A**gaacccaatcg**C** (SEQ ID NO:18)	GCGATTGGA (SEQ ID NO:46)	CGGAATGCA (SEQ ID NO:74)	GAGGCAAGCGGAATGCA (SEQ ID NO:102)			
inhA94	ggcatgaacccaatcga (SEQ ID NO:19)	TCGATTGGA (SEQ ID NO:47)	TGGAATGCA (SEQ ID NO:75)	GAGGCAAGTGGAATGCA (SEQ ID NO:103)			
inhAS94A	ggcatgaacccaatcg**C** (SEQ ID NO:20)	GCGATTGGA (SEQ ID NO:48)	CGGAATGCA (SEQ ID NO:76)	GAGGCAAGCGGAATGCA (SEQ ID NO:104)			
ahpC73	gc**C**gtcctcgaactcgtc (SEQ ID NO:21)	GACGAGT (SEQ ID NO:49)	CACCAGA (SEQ ID NO:77)	TGCATGTCACCAGA (SEQ ID NO:105)	cggcgttcagc aagctc (SEQ ID NO:118)	38	M-5167
ahpCD73H	gc**C**gtcctcgaactcgt**G** (SEQ ID NO:22)	CACGAGT (SEQ ID NO:50)	GACCAGA (SEQ ID NO:78)	TGCATGTGACCAGA (SEQ ID NO:106)			
ahpC73	gcggtcctcgaactcgtc (SEQ ID NO:23)	GACGAGT (SEQ ID NO:51)	CACCAGA (SEQ ID NO:79)	TGCATGTCACCAGA (SEQ ID NO:107)			

*ahp*CD73H	gcggtcctcgaactcgtG (SEQ ID NO:24)	CACGAGT (SEQ ID NO:52)	GACCAGA (SEQ ID NO:80)	TGCATGTGACCAGA (SEQ ID NO:108)			
*ndh*268	ggcTccgtccggcg (SEQ ID NO:25)	CGCCGA (SEQ ID NO:53)	GGCGCA (SEQ ID NO:81)	TACGAGGGCGCA (SEQ ID NO:109)	cagaccttgca ggccgactc (SEQ ID NO:119)	38	I-591
*ndh*R268H	ggcTccgtccggcA (SEQ ID NO:26)	TGCCGA (SEQ ID NO:54)	AGCGCA (SEQ ID NO:82)	TACGAGAGCGCA (SEQ ID NO:110)			
*ndh*268	ggcaccgtccggcg (SEQ ID NO:27)	CGCCGA (SEQ ID NO:55)	GGCGCA (SEQ ID NO:83)	TACGAGGGCGCA (SEQ ID NO:111)			
*ndh*R268H	ggcaccgtccggcA (SEQ ID NO:28)	TGCCGA (SEQ ID NO:56)	AGCGCA (SEQ ID NO:84)	TACGAGAGCGCA (SEQ ID NO:112)			

Shown (left to right): linear primer (LP) sequence, additional HP tail (HP), substituted linearized-HP (LHP) tail, substituted extended-LHP (ELHP) tail. Mutations (either SNPs or secondary mutations) are shown underlined in bold capitals

[1]These primers detect the SNPs as depicted in Figure 3 but using a complementary strand.

[2]Base-pairs (5'-end tail sequences not included)

[3]Not available

Example 2: Real time PCR

All PCR reactions were performed in an Applied Biosystems 7900HT sequence detector system with the 384-well block for real-time PCR. Thermal conditions were as follows: Stage 1: 95°C 10 minutes, 70°C for 30 seconds; Stage 2: 72°C for 30 seconds, 95°C 20 seconds, 69°C for 30 seconds lowering one degree in the last step for every cycle during 10 cycles; and Stage 3: 72°C for 30 seconds, 95°C 20 seconds, 60°C for 30 seconds, repeated 40 times. Data was collected in the last step of stage 3 for analysis with the SDS software version 2.0a23 (Applied Biosystems, CA). Every well was loaded with PCR cocktail containing: 1X Amplitaq Gold polymerase buffer, 0.15 U of Amplitaq Gold polymerase (Perkin-Elmer, CA), 2 mM $MgCl_2$, 2.5 pmol of each primer, 1X SYBR green I (Molecular Probes Inc., OR), 1.75 ng ROX [6-carboxy-X-rhodamine, succinimidyl ester (6-ROX, SE)] (Molecular Probes Inc., OR) (used as a reference dye), and either 0.1 ng of chromosomal DNA, or 10^5 molecules of the artificial template, or an equal volume of water (no DNA control), followed by sufficient water to result in a final volume of 5µl.

Example 3: HP assay design

Primers (Invitrogen, CA, or Illumina, CA), whose sequences are shown in Table 1, were designed using the Primer Express Software version 2.0 (Applied Biosystems, CA) to produce short amplicons (30-90 base pairs long), and to anneal between 60-65°C. A tail was added to the 5'-end of the SNP-detecting primer in order to produce a stem with the 3'-end of the primer. The stem was designed using mfold software (see bioinfo.rpi.edu/applications/mfold/old/dna/ of the world wide web) to have a Tm of 67-70°C with a free energy ΔG) between -0.5 and -2.0. Two single-stranded artificial templates (WT and MUT) were designed to test the discriminatory power of

each primer set, and chromosomal DNA of *M. tuberculosis* H37Rv was used as a WT control.

Example 4: HP High throughput assay

384-well plates (Applied Biosystems, CA) were loaded with 5 µl per well of the SNP specific and constant primer mix using a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). Plates were completely dried overnight inside a laminar flow cabinet, and kept in air-tight plastic bags at -20°C until used. Five µl/well of the PCR cocktail containing all the components except the primers were then loaded into the microtiter plates. The plates were vortexed and then centrifuged prior to being loaded into the robot of the sequence detector system apparatus.

Example 5: Comparison of linear primers (LP) and HP

The ability of LPs and HPs to distinguish between two or more SNP alleles in seven different SNP assays was compared. The principles of the HP assay are set forth in Figures 1A and 1B. For this comparison, sets of LPs for standard ARMS assays were first designed using Primer Express software. Then, a second set of primers was designed identical to the first set except that the SNP-specific primer was modified to form a stem-and-loop structure as described above. Assays containing the conventional LPs and otherwise identical assays containing the HPs were then tested for their ability distinguish between *M. tuberculosis* SNPs. Seven actual drug resistance-associated SNPs present in *M. tuberculosis* were tested, using *M. tuberculosis* chromosomal DNA to ensure that the results would be applicable in the subsequent investigations. Each assay was performed in quadruplicate and the average Ct values for each quadruplicate were calculated. Corresponding LP and HP assays were compared on two characteristics: 1) the ability to designate the correct SNP (versus an incorrect or indeterminate assignment), and 2) the average cycle threshold

difference (ΔCt) between the reactions containing primer-template matches and the reactions containing primer-template mismatches (where a larger ΔCt indicates a more robust assay). Assays were considered indeterminate if the ΔCt was lower than five. The HP assay was found to have a much greater discriminatory power than the LP assay. The LP assay identified the correct SNP in only 7 of 13 assays (producing one incorrect and five indeterminate results), while the HP assay identified the correct SNP in all 13 assays, demonstrating the superiority of this format (see Figures 2A and 2B, triangles). Furthermore, the HP assay appeared to be more sensitive for SNPs as the ΔCt values were greater in the HP assays compared to the LP assays (See Table 2).

Table 2: Ct averages of 7 SNP assays using different primer sets

Primer set	Match[2] Ct	Mismatch[3] Ct	ΔCt[4]	Rejected assays[5]
LP	25.0	32.6	7.6	6
HP	19.6	30.8	11.2	0
LHP	19.6	29.1	9.5	1
ELHP	21.1	32.1	11.0	1

[1]See Example 1 for description of isolates used
[2]Match reaction: PCR reaction in which the 3'end of the primer is complementary to the target
[3]Mismatch reaction: PCR reaction in which the 3'end of the primer is not complementary to the target
[4]Average of mismatch Ct - average of matched Ct
[5]Out of total of 13 assays performed

Example 6: HP comparison with linearized HP (LHP) and extended-LHP (ELHP).

To investigate the possibility that the enhanced ability of the HP configuration might be due to the extended 5′ tail of the SNP-specific primer rather than to the actual stem-and-loop, "linearized" HPs (LHPs) were created by mutating the 5′ tail of each HP primer so that it could no longer form a stem with the 3′ end. The LHPs were designed so that no new

secondary structures were formed and so that they had the same GC content as the HPs. Each SNP assay was repeated with LHP sets. The performance of the LHP assays was intermediate between that of the linear primers and HP primers (see Figure 2C). LHP assays only produced one indeterminate result, however the ΔCt values averaged 1.7 cycles less than the HP assays (Table 2). To determine if further lengthening of the 5′ tail would continue to enhance the ability of LHPs to distinguish among SNPs, a corresponding set of "extended" LHPs (ELHPs) was created by doubling the length of each 5′ tail. The performance of assays with ELHPs was compared with the other primers. It was found that ELHP assays produced results that were similar to LHP assays, with one indeterminate result out of 13 (see Figure 2D, triangles). Notably, assays with ELHPs had improved ΔCt values, with an average ΔCt that was equivalent to the average ΔCt of reactions containing HPs (Table 2). However, the ΔCt values of individual HP assays fell within a more narrow range than the ΔCts of the ELHPs, thus indicating that the HP-based assays remained superior.

Example 7: Evaluation of the insertion of a secondary mismatch

The effect of inserting additional base pair mismatches into the HPs was tested. A secondary mismatch was inserted into each primer either towards the center of the loop or at the end of the stem. Mismatches were designed so that total GC content was maintained. The results indicated that the secondary mismatches resulted in later Cts in both match and mismatch reactions, and led to ΔCt values that were improved by 1.1, 0.2, 2.1 and 0.8 cycles for LP, HP, LHP and ELHP respectively (Figure 2, diamonds). The number of rejected assays did not vary with the secondary mismatch, except for LPs where only four reactions were rejected (one incorrect and 3 indeterminate results). The incorporation of a secondary

mismatch can also play an important role during the HP design. The mismatch confers flexibility to the design process and can be used to avoid undesired secondary structures.

Example 8: The HP approach for loci containing multiple alleles

A single codon can contain more than two SNP alleles. This is the case for position 315 in the *katG* gene of *M. tuberculosis*, which is the most common position mutated in INH resistant clinical isolates (Ramaswamy and Musser Tuber. Lung Dis. 1998 79:3-29). The ability of the HP assay to test for four possible alleles at this position was investigated. A single WT HP primer and three different MUT HP primers were designed to be complementary to each *katG*315 allele. The HPs were then tested in assays with chromosomal *M. tuberculosis* DNA containing each mutation. The results show that the HP assay can easily distinguish among all four alleles (Figure 3).

Example 9: Success rate of large-scale HP assay design

HP assays were designed for 207 different *M. tuberculosis* SNPs at 98 different polymorphic sites previously associated with resistance to INH to test the utility of the HP approach in large-scale SNP analysis. Each assay was tested on chromosomal DNA from *M. tuberculosis* H37Rv (WT control) and both artificial templates. MUT chromosomal DNA was also used when available. Ninety-one functional HP assays (most of which included a secondary mutation) were successfully designed that detected SNPs in the *katG*, *kasA*, *ahpC*, *inhA*, *mabA* and *ndh* genes of *M. tuberculosis*. Assays that detected insertions and deletions were developed using the same parameters. Design success rates as a function of number of design attempts are shown in Table 3.

Table 3: Success rate of HP assay design

Attempt[1]	Success rate[2]
1	72.4% (71/98)
2	83.7% (82/98)
3	88.8% (87/98)
4	92.9% (91/98)

[1]Number of times a HP assay was redesigned until discriminatory

[2]Percentage of HP assays able to discriminate amongst SNP alleles according to the predefined criteria at a particular design round

The sensitivity of the assays in terms of the amount of chromosomal DNA required was also examined. Most assays gave consistent results with less than 0.05 ng/well of chromosomal DNA. Preferably 0.1 ng/well of chromosomal DNA is used as this amount resulted in smoother amplification curves.